Developing Sensor Activity Relationships for the JPL Electronic Nose Sensors Using Molecular Modeling and QSAR Techniques

A.V. Shevade, M. A. Ryan, M. L. Homer, A. D. Jewell,
H. Zhou, K. Manatt, and A.K. Kisor, A.M. Manfreda and C. J. Taylor

Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena CA 91109
Abhijit.Shevade@jpl.nasa.gov, Margaret.A.Ryan@jpl.nasa.gov, Margie.L.Homer@jpl.nasa.gov,
April.D.Jewell@jpl.nasa.gov, Hanying.Zhou@jpl.nasa.gov, Kenneth.S.Manatt@jpl.nasa.gov,
Adam.Kisor@jpl.nasa.gov, ali@manfreda.org, ctaylor@pomona.edu

***Abstract*— We report a Quantitative Structure-Activity Relationship (QSAR) study using Genetic Function Approximation (GFA) to describe the polymer-carbon composite sensor activities in the JPL Electronic Nose (ENose), when exposed to chemical vapors at parts-per-million (ppm) concentration levels. A unique QSAR molecular descriptor set developed in this work combines the default analyte property set (thermodynamic, structural etc.) with sensing film-analyte interactions that describes the sensor response. These descriptors are calculated using semi-empirical and molecular modeling tools. The QSAR training data set consists of 15-20 analyte molecules specified by NASA for applications related to Life Support and Habitation in space. The statistically validated QSAR model was also tested independently to predict the sensor activities for test analytes not considered in the training set.**

Key Words: Multivariate analysis, Molecular modeling, Quantitative Structure-Activity Relationship (QSAR), Electronic nose, Environmental monitoring, Polymer composite,

I. INTRODUCTION

The JPL electronic nose (ENose) consists of polymer-carbon composite sensing films used to detect organic vapors and other environmental contaminants for Life Support and Habitation in space applications [1,2]. The measured sensor responses are conductivity changes in polymer-carbon composite films upon exposure to target vapors (or analytes) at parts-per-million (ppm) concentrations. Training an array for a given set of analytes and a given set of environmental conditions (temperature, pressure, and humidity) is time consuming. In addition, developing training sets and calibration information may decrease the useful lifetime of the sensors. Hence the goal is to develop a sensor response model, that provides us the ability to predict sensor responses accurately and would also be of great help in characterizing sensing materials.

Previous modeling efforts [3-5] to predict the response of polymer-film based sensors, primarily on pure polymer films, use Linear Solvation Energy Relationships (LSER) and solubility parameters to obtain a good correlation between the calculated and measured responses. Modeling efforts in the past for polymer-carbon composite sensing films based on solubility parameters [6], have taken into account only the effect of polymer-analyte interactions and assumed that neither carbon nor analyte in the film play a role in sorbing analyte molecules or in contributing to the response of the film. This model of sensor response may not represent a complete picture of response in polymer-carbon composite sensors.

We report a Quantitative Structure-Activity Relationships (QSAR) study using Genetic Function Approximations (GFA) with a unique molecular descriptor set, to describe the activities of the JPL ENose polymer-carbon sensors. In our work, the sensor activity for a given analyte is defined as the coefficient, A_1, which is correlated to the sensor response as $y=A_1x + A_2x^2$, where x is the concentration of analyte and y is the measured sensor response. A_2 is generally three to five orders of magnitude smaller than A_1. The unique QSAR descriptor set combines the default analyte properties (structural, spatial, topological, conformational, and thermodynamic) with descriptors for sensing film-analyte interactions, which describe the sensor response. The modeled analyte descriptors are calculated using Quantitative Structure-Property Relationship (QSPR) techniques. The polymer-analyte interaction energies that describe sensor response are calculated using molecular modeling tools.

0-7803-9056-3/05/$20.00 © 2005 IEEE.

II. EXPERIMENTAL DETAILS

The JPL ENose array consists of 32 sensors made from 16 polymer-carbon composite films[1,2]. The JPL ENose polymers represent five categories of chemical functionality: Hydrogen-Bond Acidic, Hydrogen-Bond Basic, Dipolar and Hydrogen-Bond Basic, Moderately Dipolar and weakly H-Bond basic or acidic, and Weakly Dipolar with weak or no hydrogen-bond properties. The training of the JPL ENose sensors is done by delivering measured concentrations of analyte with controlled humidity at a given sensor substrate temperature. Response data for each analyte and each sensor are fit to an equation of the form $y=A_1x + A_2x^2$, where x is the analyte concentration and y is the normalized change in resistance. To develop sensor activity model that is concentration independent, the QSAR studies use the coefficient A_1 as the sensor activity to be correlated with molecular descriptors. The poly(styrene-co-maleic acid) sensing film is selected for this study. The choice of the current system is to further add to our previous understanding by developing a model that relates the sensor activity to molecular descriptors. The analytes and concentration ranges considered for this study are shown in TABLE I.

III. QSAR MODEL DEVELOPMENT

A. QSAR Descriptors: Analyte properties and Sensor response

The QSAR descriptors considered for the study include those that describe intrinsic analyte properties as well as sensor response.

The default analyte descriptors for QSAR studies [7] that describe the intrinsic analyte properties fall into Electronic, Spatial, Structural, Thermodynamic and Topological categories. Sensor response at a molecular level is described by interactions between the sensing film and the analyte. Descriptors for sensor response include interaction energies of the sensing film components (polymer and carbon-black) with the analyte molecules and water. Also added to the default descriptor set, is the vapor pressure of the analytes at 300K. The rationale behind using vapor pressure as a descriptor is based on functional dependence of the sensing film partition coefficient to the bulk analyte concentration.

The pair interaction energies considered to represent the sensor response descriptors are: polymer-analyte, CB-analyte, polymer-water, and CB-water, analyte-analyte and analyte-water (CB: carbon black). Depending on the type of interactions considered, the suffixes could be polymer (p), carbon black (cb), analyte (a), or water (w). For example interaction energy between the polymer (p) and the target analyte (a) is denoted by E_{pa}.

A combined descriptor set that includes the default analyte descriptors along with sensor response descriptors was used in the QSAR studies to correlate the sensor coefficients A_1 with the molecular descriptors.

B. QSAR Descriptor calculations

The default analyte descriptors were predicted by empirical and semi-empirical Quantitative Structure Property Relationships (QSPR) using the commercial software Cerius^2 [8] on a Silicon Graphics O2 workstation.

The sensing film-analyte interaction energies were calculated using Monte Carlo simulation techniques. Details on the molecular models and forcefield used for the sensing film-analyte interaction calculations are discussed elsewhere [7,9]. The interaction energy between the polymer and an analyte molecule, E_{pa}, was calculated by fixing the polymer structure in space and sampling the analyte molecule around the polymer, then averaging the energy calculated over all the random analyte configurations generated around the polymer. In this study we used 10^5 configurations. E_{pw}, $E_{p\text{-}cb}$, $E_{cb\text{-}a}$, $E_{cb\text{-}w}$, $E_{cb\text{-}cb}$, E_{aa}, and E_{aw} descriptors were calculated similarly. The BLENDS module in the Cerius^2 software performs the calculations based on the methodology described above.

TABLE I: ANALYTE LIST AND CONCENTRATION RANGE TESTED IN PARTS-PER-MILLION (PPM) FOR ENOSE OPERATION (760 TORR, 23 °C). DATA ARE TAKEN AT A CONSTANT HUMIDITY OF 5000 PPM WATER (APPROXIMATELY 18% RELATIVE HUMIDITY).

Analyte	Concentration tested *Low - High (ppm)*
1. Acetone	64 - 600
2. Ammonia	6 - 60
3. Chlorobenzene	3 - 30
4. Dichloromethane	10- 150
5. Ethanol	200-6000
6. Isopropanol	30- 400
7. Xylenes (mixed)	33- 300
8. Tetrahydrofuran	13-120
9. Trichloroethane	7- 200
10. Acetonitrile	1- 25
11. Ethylbenzene	20-180
12. Freon113	15-500
13. Hexane	15-150
14. Methyl ethyl ketone	15-150
15. Methane	1600-50000
16. Methanol	6-100
17. Toluene	5-50
18. Benzene	10-100
19. Indole	25- 450
20. Dichloroethane	10-100

C. QSAR equation: functional form and sensor activity representation

The reliability and significance of the developed QSAR models are determined by using statistical parameters such as correlation coefficient $\mathbf{r^2}$ and **F**. The **F** value is a ratio of explained and unexplained variance. As the **F** value increases, the significance of the QSAR equation becomes greater.

The first step in the QSAR model development process is to determine the number of terms (N_{term}) and the functional forms (linear or linear-quadratic or spline) to be used in the QSAR equations. The decision on N_{term} and the functional forms (linear or linear-quadratic or spline) to be used was made after evaluating the QSAR cross-validated $\mathbf{r^2}$ as a function of N_{term} for each functional form type. Based on these results [7] we have chosen to use linear-quadratic equation with 3-4 terms.

The next step is to generate a set of cross-validated equations using GFA algorithms. The response of a polymer carbon sensing film to a given analyte molecule is based on how the sensing film components (polymer and carbon black) in the polymer-carbon composite interact with the analyte molecule. Therefore, the QSAR equation that we have chosen to represent the given sensor activity is the statistically most significant one (largest $\mathbf{r^2}$ value) of the equation set containing the E_{pa} descriptor.

IV. Results and Disucussion

The polymer selected for the current investigation is poly(styrene-co-maleic acid) (Figure 1), which falls in the hydrogen bond acidic category. The pKa values at 298 K for the two carboxylic acid residues in the maleic acid portion of the co-polymer poly(styrene-co-maleic acid), on the carbon atoms labeled 1 and 2 was calculated [10] as 3.99 and 5.13, respectively.

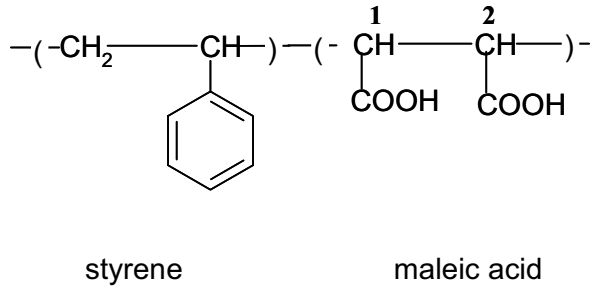

Figure 1: Poly(styrene-co-maleic acid) monomer unit

We have used a three term linear-quadratic form for the QSAR study. Training set for model development included analytes 1-17 (Table I). The cross-validated statistically most significant equation ($\mathbf{r^2}$=0.88, **F** =50.6) for the poly(styrene-co-maleic acid) containing the descriptor E_{pa} was:

$$\textbf{Calculated activity} = 1.00\text{E-}06\ VP^2 - 1.48\ E_{aw} + 0.39\ E_{pa} \quad (1)$$

The descriptor VP, refers to the vapor pressure of the analyte. The strong interaction energy between the water and the two carboxylic groups in the polymer results in dominance of the analyte-water interaction energy term, hence the appearance of the E_{aw} in equation (1). A plot of QSAR calculated versus experimental sensor activities (A1 coefficient) for the training data set is shown in Figure 2. Equation (1) was used to predict the sensor coefficient A_1 for the test analytes: benzene, dichloroethane and indole. As seen in Figure 2, the model satisfactorily predicts A_1 for benzene, dichloroethane and indole.

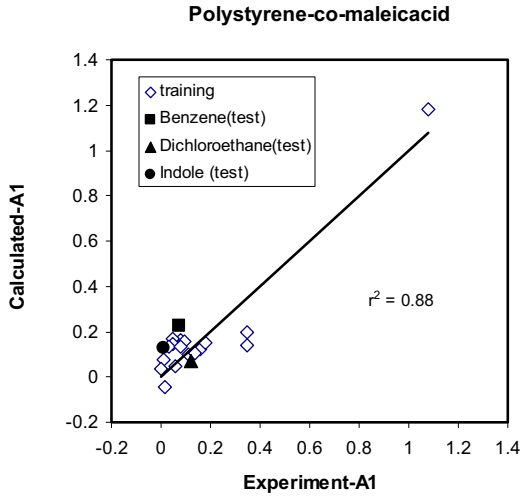

Figure 2: A plot of QSAR calculated vs. experimental sensor activity for the training and test analyte set using combined descriptor set. The calculated values for both the training and test analyte set are obtained using Equation (1). The r^2 value refers to the correlation between the calculated vs. the experimental values obtained for the training data set.

It should be noted that we do not expect to calculate the exact value of the coefficient A_1 using the QSAR approach described here. It is likely that there are effects influencing sensor response which are not accounted for in the models of interaction energy or in the physico-chemical properties of the analyte. For example, any addition of pathways for electronic conduction, such as ions in the polymer matrix, would result in a decrease in resistance, but such a decrease

would not be accounted for in either the calculated interaction energies or in the default descriptor set.

V. CONCLUSIONS

Experimental data for a polystyrene-maleic acid carbon black sensor was correlated using QSAR with intrinsic analyte properties and molecular interaction energy terms. The model developed showed good correlation for the entire analyte set as well as analyte subsets. The descriptors that predict the polymer-carbon sensor response indicate that the polymer-analyte interaction is not the only important interaction to consider. In addition to predicting sensor response, it may be possible to elucidate the sensing mechanisms using this approach. The approach will be extended to other polymer composite sensors used in the JPL ENose system.

ACKNOWLEDGMENT

The research reported in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration. The JPL ENose development program is supported by Advanced Environmental Monitoring and Control, Life Support and Habitation, ESMD, NASA. The authors thank Prof. Anton Hopfinger and Dr. David Rogers for their valuable suggestions.

REFERENCES

[1] M. A. Ryan, H. Zhou, M. G. Buehler, K. S. Manatt, V. S. Mowrey, S. P. Jackson, A. K. Kisor, A.V. Shevade and M. L. Homer, "Monitoring Space Shuttle Air Quality Using the JPL Electronic Nose," IEEE Sensors Journal, vol 4(3), pp. 337, 2004.

[2] M. A. Ryan, A. V. Shevade, H. Zhou and M. L. Homer, "A Polymer-carbon-composite sensor based ENose System," MRS Bulletin, vol 29, pp.714, 2004.

[3] J. W. Grate, S. J. Patrash, S. N. Kaganove, M. H. Abraham, B. M.Wise and N. B. Gallagher, "Inverse least-squares modeling of vapor descriptors using polymer-coated surface acoustic wave sensor array responses," Anal. Chem., vol. 73 (21), pp. 5247, 2001.

[4] A. Hierlemann, E. T. Zellers and A. J. Ricco, "Use of linear salvation energy relationships for modeling responses from polymer-coated acoustic-wave vapor sensors," Anal. Chem., vol. 73 (14), pp. 3458, 2001.

[5] M. P. Eastman, R. C. Hughes, G. Yelton, A. J. Ricco, S. V. Patel, M. W. Jenkins, J. Electrochem. Soc., vol. 146 (10), pp. 3907, 1999.

[6] W.A. Goddard III, T. Cagin, M. Blanco, N. Vaidehi, S. Dasgupta, W. Floriano, M. Belmares, J. Kua, G. Zamanakos, S. Kashihara, M. Iotov and G.H. Gao, "Strategies for multiscale modeling and simulation of organic materials: polymers and biopolymers," Comput. Theo. Poly. Sci., vol. 11 (5), pp. 329, 2001.

[7] A.V.Shevade, M. A. Ryan, M. L. Homer, A. D. Jewell, H. Zhou, K. Manatt, and A.K. Kisor, "Correlating Polymer-Carbon Composite Sensor Response with Molecular Descriptors, submitted to the Journal of Electrochemical Society.

[8] Cerius^2 v 4.2, Accelrys Inc., San Diego, California, USA (2000)..

[9] A. V. Shevade, M. A. Ryan, M. L. Homer, A. M. Manfreda, H. Zhou and K. Manatt, Sens. and Actuat. B: Chemical, 93, 84 (2003).

[10] pKa program, Advanced Chemistry Development (ACD)